DOCUMENT RESUME

ED 269 434 TM 860 264

AUTHOR Khammash, Salma B.
TITLE A Pictorial Test Assessing Aural-Vocabulary of Arabic-Speaking Children at Kindergarten and Early Primary Levels.
PUB DATE 17 Apr 86
NOTE 18p.; Paper presented at the Annual Meeting of the American Educational Research Association (70th, San Francisco, CA, April 16-20, 1986).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Arabs; Auditory Stimuli; Early Childhood Education; Foreign Countries; *Foreign Students; Individual Testing; Language Proficiency; *Limited English Speaking; Norm Referenced Tests; Pictorial Stimuli; *Test Construction; Test Reliability; Test Validity; *Visual Measures
IDENTIFIERS *Arabic Picture Vocabulary Test; *Jordan; Peabody Picture Vocabulary Test (Revised)

ABSTRACT

This paper presents the development of the Arabic Picture Vocabulary Test (APVT), derived from the Peabody Picture Vocabulary Test-Revised, for use in assessing the aural-vocabulary of newly arrived limited English proficiency children, from kindergarten through grade three, whose first language is Arabic. The APVT, an individually administered norm-referenced power test was designed for students who have normal sight and hearing, and who understand Modern Standard Arabic to some degree. The 150 item test, arranged in order of increasing difficulty, consists of a page of pictures which together with a spoken word constitutes a test item. The task is to identify a picture that best illustrates the stimulus word meaning. Testing requires 20 to 30 minutes and is scored during administration. The APVT, translated and adapted to Arabic, was validated and standardized in Jordan on a representative sample of 2,900 students (1,450 girls, 1,450 boys). Raw scores were converted to age-referenced norms. Test-retest reliabilities were considered to be an unbiased estimate of reliabilities of the corresponding standard score equivalents. Field testing of the APVT is currently underway in the Dearborn Public Schools (Michigan), in an attempt to cross validate the APVT. (Author/PN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED269434

A PICTORIAL TEST ASSESSING AURAL-VOCABULARY OF ARABIC-SPEAKING CHILDREN AT KINDERGARTEN AND EARLY PRIMARY LEVELS

Salma B. Khammash
The University of Michigan

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
S. B. Khammash

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Paper presented at the American Educational Research Annual Meeting, AERA/SIG International Studies San Francisco, April 17, 1986.

TM 860 264

ABSTRACT

A PICTORIAL TEST ASSESSING AURAL-VOCABULARY
OF ARABIC-SPEAKING CHILDREN AT KINDERGARTEN
AND EARLY PRIMARY LEVELS

Salma B. Khammash

This paper discusses the methods and procedures followed in developing the Arabic Picture Vocabulary Test (APVT). The APVT, derived from the Peabody Picture Vocabulary Test-Revised (PPVT-R), was designed to assess levels of aural-vocabulary (for proper grade placement) of incoming Arabic-speaking children with limited English proficiency (LEP). Their scores can be more generally interpreted as measures of verbal ability, in so far as they assess, through hierarchy of vocabulary difficulty, the level of semantic complexity at which the examinee is able to understand words.

The APVT, an individually administered norm-referenced power test, was devised for students ranging in age from 5 yesrs to 9+, who have normal sight and hearing, and understand Modern Standard Arabic to some degree. The instrument contains one hundred and fifty test items (preceded by seven prectice items), arranged in order of increasing difficulty. A page of pictures together with a spoken word constitutes a test item. The child's task is to identify the picture considered to illustrate best the meaning of a stimulus word presented orally by the examiner, and as the test proceeds, a point is reached where the items cease to be fully comprehended. Test administration is simple and can be undertaken by any experienced teacher or school official who follows instructions carefully. However, the examiner must be fluent in Arabic in order to read the directions to the student. Testing requires 20 to 30 minutes, because the examinee must answer only those items of suitable difficulty. Scoring, which is objective and rapid, is accomplished while the test is being administered.

The APVT, translated and adapted to Arabic by this writer, was validated and standardized in Jordan on a carefully selected (representative) sample of 2900 students (1450 girls, 1450 boys), grades K-6 for test validation, and K-3 for standardization. Raw scores were converted to age-referenced norms.

A PICTORIAL TEST ASSESSING AURAL-VOCABULARY OF ARABIC-SPEAKING CHILDREN AT KINDERGARTEN AND EARLY PRIMARY LEVELS

Salma B. Khammash

This paper presents the development of the Arabic Picture Vocabulary Test (APVT), derived from the Peabody Picture Vocabulary Test-Revised (PPVT-R), for use in assessing the aural-vocabulary of newly arrived limited English proficiency (LEP) children, grades K-3, whose home language is Arabic. However, the primary purpose of the paper is not to focus on the specific test instrument, but to attend to the processes of: 1) translating and adapting the PPVT-R to Arabic (language and culture), 2) the procedures followed in the standardization and norm development of the APVT, and 3) a presentation of evidence of reliability estimate (coefficient of stability) of the developed instrument.

The processes and outcomes described in this paper are proposed as generalizable in the development of other tests in the same language (Arabic) or other languages and cultures.

The Need: The research reported here was undertaken to meet the need for a psychometrically sound testing instrument to ascertain, to the degree feasible, the cognitive level of limited English proficiency (LEP) children whose home language is Arabic. There is a large number of these children entering American schools, among them many who arrive in the U.S. with few or no documents. School officials often have little beyond the physical maturity of the child to guide them in making the proper class placement.

The administrator is often compelled to make such a crucial decision as that of class or grade placement while completely lacking solid basis for selection. If the administrator (principal) or a staff member is a speaker of Arabic an interview can be conducted with the student and appropriate placement is made. However, without personnel able to communicate with the student in his/her home language an inappropriate placement may be made.

Early in the course of the Arabic Language Bilingual Materials Development Project a particularly urgent need was identified. It was for a testing

instrument to assist in measuring the stage of cognitive development of the monolingual Arabic-speaking child. None was found to exist in Arabic. This writer, as the Director of Research and Evaluation of the Project, assumed the task of finding a solution to this problem.

After an in-depth survey of available instruments which might serve as a point of departure -- one which might be translated and adapted to the Arabic language and culture -- the Peabody Picture Vocabulary Test was selected for this purpose (permission was secured from the author, Dr. L. Dunn). In addition, if a translated version of a wide range test and most commonly used measure of receptive vocabulary such as the PPVT is to be useful, then the basal and ceiling point concept must be operative in an adapted version. The basal-ceiling concept presupposes a near perfect correlation between item order and item difficulty.

A pilot study was carried out in the fall of 1980. Seventy-five items from each Form, A & B, of the PPVT were translated to Arabic (Modern Standard Arabic), the distractors were selected, reorganized and numbered according to the Arabic system -- right to left. The resulting Arabic test was administered individually by Arabic-speaking bilingual teachers to 30 newly arrived children of both sexes, ages 5 to 8 years, in the Dearborn Public Schools, Dearborn, Michigan. Fifteen children took Form A, and 15 students took Form B of the PPVT. Although the pilot sample was small and the reliability figures were imprecise, it appeared that the test was usable; consequently revisions were made to: a) select the proper Arabic stimulus words, b) select and adapt optional responses suited to the Arabic language and culture, and c) add more items to provide a sufficient range of difficulty.

The results of the pilot study indicated that a practical test of aural-vocabulary in Arabic could be developed by translating and adapting the 1965 edition of the PPVT. However, in the meantime, a revised version of the test, the PPVT-R, was published in early 1981. In developing the PPVT-R, the authors (Lloyd & Leota Dunn) selected and validated items using a national sample of students with adequate representation of minority groups. As a result, the revised edition should be less subject to cultural bias than the original test. For this reason, the PPVT-R was chosen for the comprehensive translation and adaptation project.

METHODS

Selection of Items

The first step was to examine the 350 stimulus word pool of the PPVT-R (including both Forms L&M) for which illustrative drawings were available. After thorough examination of the item pool, it was decided that an adequate sample of words could be drawn from this source to represent the levels of Arabic-speaking children between the ages of 5:0 and 12:0. It was further decided that provided care were taken to eliminate incongruities and to match pictures with words in accordance with Arabic linguistic usage and cultural associations, the use of some of the PPVT-R drawings for Arabic-speaking children would not invalidate inferred correspondences between words and pictures. Since it had been agreed to carry out field testing with Arabic-speaking children it was considered that any failure on the part of the writer in the initial selection of items would be detected during item analysis and subsequent validation.

From the 175 items per Form (L&M), 110 stimulus words and the associated response options which had displayed linear growth curves, and which were considered amenable to adaptation were identified as suitable for this purpose. To ensure comparability between Form L and Form M item pools, selections were balanced with respect to the content categories in the PPVT-R.

Translation and Adaptation of Items

For some items there was one-to-one correspondence between the English and Arabic experience with the object pictures on the plate (e.g., human body parts). In this case all that was necessary was to substitute the appropriate Arabic term for each object, and to rearrange the items in the standard Arabic format. However, there were other items, such as those dealing with musical instruments, in which the items had to be adapted by choosing from the Arab culture objects of equivalent familiarity and function to their counterparts in the English version. For example, the 'Ud (lute), a musical instrument used at almost every Arab celebration was substituted for Guitar, or the Flute for Clarinet in the English version.

To maintain intra-item comparability of optional responses, the criteria used to develop the original plates were followed in producing the Arabic

version. These included: 1) the four plate drawings, all represented objects familiar to the age group for whom they were intended, 2) the drawings were of equal size, 3) the drawings were of equal intensity and detail, and 4) there was male-female balance on plates containing humans.

The translated and adapted items were grouped into two prototype tests comprised of 110 stimulus items each from Forms L and M respectively. One hundred and ten plates were made up, each of four pictures, to provide correspondences and functional distractors for the stimulus words. Two books of plates (Forms L and M) were prepared, each containing 110 plates arranged in ascending order of anticipated difficulty. Sufficient copies and the accompanying individual test records for testing a large sample of students were also prepared in Arabic.

Sample Characteristics

The resulting tests (Forms L&M) were tried out under standard conditions by administering them to a representative sample of Arabic-speaking students in Jordan in the spring of 1981. Jordan was chosen as the testing site was because of the large population of children in the United States who come from the general demographic area of Jordan and its neighbors, where the speech (terminology and vocabulary closest to Modern Standard Arabic) would fit the experience of the LEP Arab child in U.S. public schools. Also, Jordan was accessible to the project, and a suitably large variety of Arab communities and their respective dialects of Arabic are present (Egyptian, Iraqi, Lebanese, Palestininan, Syrian, Yemeni, etc.). Basically, we wished to determine whether the translated and adapted test items adequately discriminated among Arab students at each of the targeted age levels. In operational terms, an item was considered adequate for a particular age level if between 40 and 60 percent of the students in the group answered the item correctly.

The sampling plan was based on population data from the 1973 Census and the Jordanian Ministry of Education - Educational Survey 1980. The stratification criteria used to ensure adequate representation included: (1) grade and gender balance, (2) appropriate geographic, school and socioeconomic representation, and (3) number tested by form.

Student Sample. The validation sample consisted of 1400 students from grades K through 6. A total of 200 students were included within each of the seven grade groups. The sampling plan specified that equal numbers of females and males be tested within each grade group; 100 females and 100 males were

selected at random at each grade level. Student age ranged from 5 to 12, with approximately even distribution across grades at one-year intervals.

Ideally, we would have sampled an equal number at each age level, but this was not feasible due to the involved governmental permission process. However, the school and and grade sampling process did yield age groups sufficiently large to compute reasonably stable data, and a fairly close balance between female and male students. Distribution of the validation sample by age and sex appears in Table 1, p. 12.

Schools and Geographic Locations. To ensure adequate representation of the various Arab communities, language dialects and socioeconomic status, the sample was selected across a wide variety of schools and geographic locations in Jordan.

Three geographic regions in addition to Amman, the largest city and capital of Jordan, were used in the validation activities - North, South and Center of the country. Students from 55 schools throughout Jordan participated in the testing program. Since schools beyond the kindergarten are segregated by sex, complementary girls' and boys' schools were selected in each community. Also, the majority (74%) of Jordanian elementary students attend government (public) schools; therefore, more government schools were selected than private schools. In addition, an attempt was made to include students from UNRWA (United Nations Relief Works Agency) schools. The selected sites included 43 schools (28 government, 13 private, and 2 UNRWA) located in and around Amman, and 12 schools, all government, located in towns and villages in the North, South and Center of Jordan. Overall, 960 students (69%) were selected from government elementary schools, and 440 students, including kindergarten, from private and UNRWA elementary schools.

Test Administration. Since there were two prototype Forms (L&M) of the test in Arabic, test form (L or M) was randomly assigned to examinees (only one form was administered to each examinee). Test forms were alternated so as to split the total sample into equal halves. As a result, a total of 100 students at each grade -- 50 girls and 50 boys -- took each prototype form of the test. In all cases, trained examiners (12 women, 8 men) tested individually each selected examinee.

This author personally trained the examiners and supervised on-site testing activities in 30 of the 55 schools where testing took place. The testing activity was carried out with the cooperation of the Jordanian Ministry of Education.

After the test validation activities were completed in Jordan, item analysis for difficulty, discrimination and wrong answer distraction were carried out for each year-level group separately for both Forms (L&M). The item analysis provided the following information for each item: 1) item difficulty, expressed as the percentage of examinees passing an item in successive age groups; 2) item discrimination, expressed for each group as the point-biserial correlation of the item with the total score; 3) item decoy action, expressed as the percentage of examinees selecting each distractor on the plate; 4) negatively discriminating distractor information, determined by comparing the percentage of examinees in high and low ability groups who selected each distractor; and 5) sex differences, expressed as the statistical significance between the proportion of female and male examinees answering each item correctly. In addition to the above information, examiners' comments on the items were reviewed. Items were checked again to identify sex stereotypes or cultural illustrations. Furthermore, an assessment was also made of the kinds of words (nouns, gerunds for action, descriptors) selected so that a better balance across the categories in the test could be achieved.

Construction of the Arabic Picture Vocabulary Test (APVT).

Items were selected for inclusion in the final APVT if they satisfied the following criteria: (1) within the 40% to 60% difficulty range of the item, at least one of the phi* values should reach the 1% significance level; (2) outside the 40-60% the remaining phi values should be positive, except in the case of items with a sharp difficulty rise with age, where for the youngest age group it may be negative; (3) the difficulty of an item should decrease with age uniformly; (4) the phi values of each of the three distractors in the item should be negative or zero with the 40% - 60% difficulty range of the item. (5) the remaining values for the distractors should also be mainly negative, except in the case of items with steep difficulty gradients, (where these distractors may have positive phi values for the youngest age group);

---

* Guilford, J.P. Psychometric Methods, 2nd Ed. New York: McGraw Hill, 1954

and (6) a distractor was also retained if it absorbed at least 10% of the total distraction.

On the basis of these criteria, a number of items had been flagged for elimination. It was therefore decided to pool the items from both Forms L&M to make a single test. This was accomplished by grouping the remaining items for each age level, covering the age range 5:0 to 9:0. Where there were several items at a particular level, the best discriminators were chosen; and where competing items were of equal quality the required number were chosen randomly. In its final version, the consolidated APVT contained 150 items in a single book. All the age ranges covered were represented within the test by items with difficulty levels between 40% and 60% for each year group, while the upper and lower reaches of the test were covered by items with increasingly extreme difficulty levels -- which retained significant phi values. Seven practice items were also included, four for ages below eight years, and 3 for examinees 8:0 and above. The complementary individual test record was also developed by compiling the corresponding words in Arabic.

Standardization of the APVT in Jordan

The APVT was standardized in the spring of 1982 in Jordan, one year after the validation program described above, again with the cooperation of the Ministry of Education. The sampling plan used in the validation program, was followed for the standardization activities, namely the stratification criteria ensuring adequate representation of: 1) grade and sex balance, 2) geographic and school representation, and 3) appropriate socioeconomic representation.

A representative sample of 1500 students, grades K-3, stratified by sex and grade were randomly selected from 80 Jordanian schools throughout the country. The selected schools were chosen to provide a 50 percent overlap with the schools which participated in the validation study. The distribution of the standardization sample by grade and sex appears in Table 2, p. 12.

The plan was to develop norms within age levels at six-month intervals. Normally, we would have sampled an equal number at each age level. However, as explained earlier, this was not feasible due to the involved governmental permission process. Again, the school sampling process below did yield age groups sufficiently large to compute reasonably stable norms (N=100), and a fairly close balance between girls and boys, as shown in Table 3, p. 13.

Schools were selected to be representative of their geographic locations and type (government, private, UNRWA). Overall, there were 52 government (public) schools, 16 private schools, including kindergarten* schools, and 12 UNRWA schools, for a total of 80 schools. In all cases, except for the kindergarten, equal numbers of girls' and boys' schools were selected. The geographic distribution was more representative than for the validation sample. Again, 50 schools including the 6 kindergarten schools were located in and around Amman, the capital city of Jordan, (24 government, 10 private, and 10 UNRWA elementary schools). The remaining 30 elementary schools (28 government, 2 UNRWA) were located in towns and villages outside Amman -- in the North, South and South-west, and Center of Jordan. Table 4 (p. 13) contains summary information on the geographic and school-location, and students included in the APVT standardization sample.

The sample was also designed to provide a broad socioeconomic representation of Jordan. Occupation of the major wage earner in the family was used to ensure appropriate representation in the standardization sample. The sample contained approximately proportionate numbers of the occupations cited by the Jordan Department of Statistics: 1) professionals (doctor, lawyer, professor), 2) government civil employees - including teachers in government schools, 3) military service, 4) business (owner, manager, clerk), 5) skilled workers - craftsmen and kindred workers, and 6) land owners, farmers, and farm laborers.

APVT Test Procedures. After securing the permission of the Jordanian Minister of Education and the Secretary General of Education, arrangements were made for the present writer to meet with officials in the Ministry of Education in order to select examiners and a coordinator. All of the examiners selected were college graduates and at one time had taught elementary grades. Furthermore, about half of these examiners were experienced because they had been members of the team that participated in the validation process, the previous year. This author conducted the training sessions for the examiners. Meanwhile, the coordinator aided in the selection of the schools to be used as the testing sites in the standardization program.

---

* All kindergarten schools are private as there were no government kindergartens in Jordan at the time the data were gathered.

After the training period was concluded, each examiner was assigned grades and schools in various locations in Jordan. At each school, students were randomly selected within the predetermined sample categories of grade, gender and socioeconomic representation.

In addition to the standardization sample, a number of students who took the APVT in its Forms L and M in the validation program, the previous year, were also tested (by this writer) for reliability purposes.

Norms Development

Norms development began with a careful check of the individual record forms to verify the correct determination of the basal and ceiling levels. Each record form was then checked to verify the raw score entered by the examiner in Arabic. The corrected data were then entered into the computer for the computation of norms.

Deriving the Deviation-Type Norms. The following steps were used to obtain the deviation-type norms (standard score equivalents, percentile equivalents, and stanines): (1) Means and standard deviations were computed separately for each of the nine age groups tested -- six-month intervals. (2) For each of the nine age groups, the raw scores were then converted to normalized z scores based on the normal probability curve. (3) Percentile equivalents corresponding to each z score were read from a table of the normal curve. (4) Stanines were computed from the normalized z scores by the usual formula. (5) The normalized z scores were converted to unsmoothed normalized Standard Score Equivalents having a mean of 100 and a standard deviation of 15. (6) To smooth these unsmoothed distributions, the raw score data for each of the nine age groups were plotted on cross-section paper and a smoothed curve was fitted to each set of plotted points. A master table of standard score equivalents was prepared by reading from the smoothed norm plots.

RESULTS

Reliability. Reliability coefficients were computed by correlating scores on the prototype Forms (L&M) used in the validation program with scores on the APVT during the norming activities. Essentially, this provided a measure of test-retest stability over a one-year interval. As mentioned earlier, this was accomplished by systematically testing a student-sample from those who participated in the validation study the previous year.

The reliability sample of 300 examinees equalled 20% of the standardization sample. Of those, 150 students (75 girls, 75 boys) took Form L during the validation program, while the other 150 students (75 girls, 75 boys) took Form M. To provide equal distribution, 100 students participated from each of three grades: K, 1, and 2. All 300 examinees were tested with the composite APVT during the norming program conducted one year later when the students were in grades 1, 2, and 3 respectively.

The test-retest reliability of the APVT at each grade level was calculated in two steps. First, the correlations of APVT scores with each Form of the validation test were computed. However, since the APVT contained 150 items and each of the prototype forms used in the validation contained only 110 items, these correlations would systematically underestimate the test-retest correlations which would have been obtained by administering the entire APVT at both times. Therefore, the second step was to use the Spearman-Brown prophecy formula (for test length) to estimate the test-retest reliability of the 150 item APVT form. The obtained correlations ranged from .90 to .93, and the corresponding test-retest reliabilities ranged from .93 to .945.

Standard Error of Measurement. Since test interpretation was to be made on the basis of standard score equivalents, it was desirable to have the standard error of measurement of those scores. As the score distributions of the various age-groupings did not differ significantly from the normal distribution, the test-retest reliabilities were considered an unbiased estimate of the reliabilities of the corresponding standard score equivalents. Using these reliability estimates, and remembering that all standard score equivalents have standard deviations of 15, the standard error of measurement (SEM) was computed for each grouping in the reliability sample. The standard errors of measurement for the standard score equivalents, and reliabilities at each grade level appear in Table 5, p. 14.

The observed values for the SEM were remarkably consistent, and all below 4.0. It therefore seems reasonable that, until cross validation studies can be compiled, 4.0 can be used as a conservative estimate of the standard error of measurement (SEM) of the standard score equivalents at all age levels not directly tested. Since the rounded value is more convenient for practical use, SEM=4 is used to interpret individual scores at all age/grade levels.

DISCUSSION

The Arabic Picture Vocabulary Test (APVT), derived from the PPVT-R, was validated and standardized in Jordan. The choice of Jordan to carry out the testing activities was made because of the large population of children in the U.S. from the general geographic areas of Jordan and its neighbors, where the speech would fit the experience of the LEP Arab child. With this in mind, the construct validity of the APVT for assessing aural-vocabulary in Modern Standard Arabic was established by the validation procedures described in the body of this paper.

The findings of this study show that the test-retest reliabilities of the APVT as highly significant, ranging from .93 to .945. Furthermore, it was observed that the score distribution of various age-groupings in the research did not differ significantly from a normal distribution; the test-retest reliabiiities, therefore, were considered to be an unbiased estimate of reliabilities of the corresponding standard score equivalents (SEM=4). These results indicate that adapting the PPVT-R to Arabic (language and culture), although expensive and time consuming procedures, had yielded significantly higher reliabilities than simply translating the test from English to Arabic. The linguistic and cultural contrasts between the two peoples concerned presented such a formidable challenge that reconceptualizing the verbal inventory of the PPVT-R proved necessary.

The present state of knowledge in both the linguistic and the broader cultural areas is insufficient to proceed, on a comfortable basis, to firm general conclusions. It is easier to arrive at partial statements which one may make with some confidence, namely the observed usefulness of this experimental instrument. Only through cross validation carried out with sufficient population(s) over several years, will there be basis for a definite statement as to the potential usefulness of this instrument to educators in the U.S., Jordan, the Middle East or elsewhere. At this time, field testing of the APVT is underway in the Dearborn Public Schools, Dearborn, Michigan.

This activity was carried out taking into account the differences of two cultures with the cooperation of the Ministry of Education of Jordan. However, there is reason to believe that these research activities succeeded in demonstrating the efficacy of the procedures employed in this project, and give confidence that these procedures are generalizable in developing other test in the same language (Arabic), or other languages and cultures.

Table 1

APVT- Validation Sample
According to Age and Sex

(N=1400)

Age Groups	Female	Male	Total
5 years	102	102	204
6 "	122	120	242
7 "	120	120	240
8 "	92	93	185
9 "	85	83	168
10 "	81	77	158
11 "	72	73	145
12+ "	26	32	58
Total	700	700	1400

Table 2

APVT - Standardization Sample
Accoring to Grades and Sex

(N=1500)

Grades	Female	Male	Total
Kindergarten	150	150	300
First Grade	200	200	400
Second Grade	200	200	400
Third Grade	200	200	400
Total	750	750	1500

Table 3

APVT - Standardization Sample
According to Age and Sex

(N=1500)

Age Groups (yrs. & mons.)	Female	Male	Total
5-0 - 5-5	57	59	116
5-6 - 5-11	62	68	130
6-0 - 6-5	57	51	108
6-6 - 6-11	118	104	222
7-0 - 7-5	94	100	194
7-6 - 7-11	100	102	202
8-0 - 8-5	105	103	208
8-6 - 8-11	99	110	209
9-0 - 9-5	58	53	111
Total	750	750	1500

Table 4

APVT - Standardization Sample
According to Schools, Geographic Locations and Sex
in Jordan

(N=1500)

School No. and Location	Female	Male	Total
Amman and adjacent areas (50 schools)*	515	515	1030
North (6 schools)	60	60	120
South & South-West (10 schools)	80	80	160
Center (14 schools)	95	95	190
Total	750	750	1500

* Government (24), private (16), UNRWA (10)

Table 5

Standard Errors of Measurement for Standard Score Equivalents for Each Reliability Group

(N=300)

Grade At APVT Test	Pre-Test Form	Reliability Estimate	Standard Error of Measurement
Kidergarten to Grade 1	L	.93	3.97
Kindergarten to Grade 1	M	.93	3.97
Grade 1 to Grade 2	L	.94	3.69
Grade 1 to Grade 2	M	.94	3.69
Grade 2 to Grade 3	L	.945	3.52
Grade 2 to Grade 3	M	.93	3.97

REFERENCES

American Psychological Association. 1974. Standards for Educational and Psychological Tests. Washington, D.C.: American Psychological Association.

Ba'albaki, M. 1979. Al-Mawrid - A Modern English-Arabic Dictionary. Beirut, Lebanon: Dar Al-ilm Lil-Malayen.

Brown, F.G. 1983. Principles of Educational and Psychological Testing (3rd ed.). New York: Holt, Rinehart and Winston.

Dunn, L.M. 1965. Peabody Picture Vocabulary Test Expanded Manual. Circle Pines, MN.: American Guidance Service.

Dunn, L.M. and L.M. Dunn. 1981. Peabody Picture Vocabulary Test-Revised Manual. Circle Pines, MN.: American Guidance Service.

Guilford, J.P. 1954. Psychometric Methods (2nd ed.). New York: McGraw-Hill.

Jordan Department of Statistics. 1973. A Statistical Guide to Jordan. Amman, Jordan.

Jordan Ministry of Education. 1980. A Guide to Educational Surveys in Jordan for School-year 1980-81. Amman, Jordan: Survey and Statistical Division (Arabic).

Mehrens, W.A., and I.J. Lehmann. 1980. Standardized Tests in Education (3rd ed.). New York: Holt, Rinehart and Winston.

Thorndike, R.L. (ed.). 1971. Educational Measurement (2nd ed.). Washington, D.C.: American Council on Education.